DOCUMENT RESUME

ED 205 146 HE 014 215

AUTHOR Ahrens, Steve
TITLE Analysis and Classification of Entering Freshmen Mathematic Students Using Multiple Discriminate Function Analysis.
PUB DATE 80
NOTE 14p.: Paper presented at the Annual Meeting of the Southern Association of Institutional Research (Louisville, KY, Fall 1980).

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Academic Achievement: Academic Aptitude: *College Freshmen: *College Mathematics: Discriminant Analysis: Educational Background: Grades (Scholastic): Higher Education: Knowledge Level: Longitudinal Studies: Mathematics Instruction: Predictive Measurement: *Predictor Variables: *Screening Tests: *Student Placement: Testing
IDENTIFIERS *West Virginia University

ABSTRACT

Predictor variables that could be used effectively to place entering freshmen mathematics students into courses of instruction in mathematics were investigated at West Virginia University. Multiple discriminant analysis was used with nearly 6,000 student records collected over a three-year period, and a series of predictive equations were generated and later cross-validated. Beginning in August 1977 all incoming freshmen students who indicated an interest in mathematics in their registration packet were administered a one-hour timed departmental exam in mathematics. The exam consisted of 36 multiple choice questions divided into three sections of 12 questions each. The first section tested for knowledge of basic algebra: the second section tested for knowledge of trigonometry and analytic geometry: and the third section tested for knowledge of solid analytic geometry and the calculus of a single variable. At the end of the fall 1977 semester, grades for the students in the various mathematics courses were obtained and matched to the earlier individual test scores. Additionally the students' high school backgrounds in mathematics and the recommendations of the students' advisor were obtained, along with American College Testing program scores. Even before completion of the study, the prediction equations appeared to discriminate among the groups effectively (advisor recommendations were dropped from the analysis). The two best predictor variables were high school background and the grade on a departmental examination. (SW)

****
* Reproductions supplied by EDRS are the best that can be made
* from the original document.
****

ED205146

Analysis and Classification of Entering Freshmen Mathematic Students using Multiple Discriminate Function Analysis

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Steve Ahrens

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☐ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

Paper Presented at
The annual Conference of the
Southern Association of Institutional Research

Louisville, Kentucky

Fall, 1980

Steve Ahrens, Associate Director
Institutional Research
West Virginia University
209 Purinton House
Morgantown, West Virginia 26506

HE014215

ABSTRACT

During the summer of 1977 the mathematics department of West Virginia University determined that the current method of placing entering freshmen into mathematical courses was inadequate and due more to happenstance than to any valid mehtodology. As a result, this study was undertaken in conjunction with the Office of Institutional Research to determine whether a set of predictor variables could be used effectively to place entering freshmen mathematic students into courses of instruction in mathematics.

Multiple discriminate analysis was used on nearly six thousand different student records collected over a three year period and a series of predication equations were generated and later cross-validated.

This study is still an on-going one and much further enhancements are anticipated. However, the prediction equations that were developed appear to discriminate among the groups effectively. The two predictor variables accounting for the best discriminating effect were high school background and the grade on a departmentally concieved examination.

An Analysis and Classification of Entering Freshment Mathematic Students Using Multiple Discriminant Function Analysis

During the summer of 1977 the mathematics department of West Virginia University determined that the current method of placing entering freshmen into mathematical course was inadequate and due more to happenstance than to any valid methodology. As a result, this study was undertaken in conjunction with the Office of Institutional Research to determine whether a set of predictor variables could be used effectively to place entering freshmen mathematic students into courses of instruction in mathematics.

The results of this study are hereby presented and represent data gleaned over a three year period and encompass nearly six thousand individual students. For the purpose of this study, all statistical effects were considered to be fixed and inferences to students outside of West Virginia University were not considered.

METHODOLOGY

Beginning in August of 1977 all incoming freshmen students that indicated an interest in mathematics in their registration packet were administered a one hour timed departmental exam in mathematics. The exam consisted of thirty-six multiple choice questions divided into three sections of twelve questions each. The first section tested for knowledge of basic algebra; the second section tested for knowledge of trigonometry and analytic geometry; whereas, the third section tested for knowledge of solid analytic geometry and the claculus of a single variable.

HE 014 215

The completed exams were then read through an optical scanner and a subsequent data set was created and file punched into computer cards. At the end of the ensuing fall semester, the graded enrollment course files were accessed and grades for the students in the various mathematics courses were obtained and matched back to the individual test scores on social security number. In a similar manner, ACT tapes were accessed and data for ACT composite and mathematics were obtained and matched. Finally, the student's high school background in mathematics and the recommendation of the student's advisor were obtained from the Admissions and Records database maintained by the West Virginia Board of Regents. The different data elements were amalgamated onto one standard label computer tape in the form seen in Figure 1.

Figure 1

File Layout of Raw Data

Variable	Field Location	Type Data
Social Security	1 - 9	numeric
Test Score	10 - 15	numeric
HS Background	16 - 18	numeric
ACT composite	48 - 49	numeric
ACT mathematics	50 - 51	numeric
Course	52 - 53	alphanumeric
Course grade	54 - 54	alphanumeric
Recommendation	55 - 56	alphanumeric

ANALYSIS

The criterion groups used in the analysis were the individual courses that were possible to assign to the students. The course descriptions are found in Appendix A, however briefly they were as follows:

Group1 - Beginning Algebra
Group2 - College Algebra
Group3 - Plane Trigonometry
Group4 - Pre-calculus mathematics
Group5 - Introductory Calculus

The predictor variables were as such: high school background as measured by number of credits taken in mathematics; departmental exam scores with a range from zero to thirty-six; ACT scores on both the composite and mathematical portions; and, finally, course recommendation of the student's advisor. The course recommendation was later dropped from the analysis, as it proved unreliable and at various times was not collected on the database files.

To distinguish between the criterion groups, multiple discriminate analysis was performed and the resulting functions were in the form

$$D_i = d_{i1} Z_1 + d_{i2} Z_2 + \ldots + d_{ip} Z_p$$

where D_i is the value for the discriminant function i, the d's are weighting coefficients, and the $Z's$ are the standardized values of the p discriminating variables. It is noteworty in multiple discriminant function analysis that the term *multiple* is used to denote that the criterion variable has more than two levels and it does not related to predictor variables (Wilks,1932).

The first step in the analysis involved a determination of whether the various predictor variables could , as a set, differentiate between the criterion groups. This was done by testing the mean vectors of the criterion groups (see Table 1). The null hypothesis tested for was the the mean vectors were identical. The statistical test for the equality of the group centroids conventionally was measured by the Wilks' lambda statistic. Wilks' lamba was calculated to be .03168 which was equivalent to an F ratio of 24.69. The probability of obtaining an F this large by chance was less than .01 and it was concluded that the predictor variables did indeed discriminate between the groups given that the sample was representative.

Table 1

Group means by Predictor Variables

	CRITERION GROUPS				
PREDICTORS	Group1	Group2	Group3	Group4	Group5
HS Background	1.2	1.8	2.3	3.4	4.6
ACT mathematics	15.2	17.8	26.3	27.4	31.4
ACT composite	16.3	20.8	26.2	26.4	29.8
Exam Score	8.2	11.4	15.6	19.8	24.8

As the it was determined that the predictors did discriminate between groups, the prediction equations were generated. Table 2 contains the coefficients of the four variables in the discriminant equation. The normalized vectors are the coefficients of the variables in the discriminant equations. The scaled vectors show the relative contribution of each variable to group discrimination; indicating the relative importance of each of the predictors in construction of the discriminant function.

Table 2

Normalized and Scaled Vectors for Discriminant Functions I and II for the Four Predictor Variables

	VECTORS			
	Normalized		Scaled	
Predictor Variables	I	II	I	II
HS Background	.7230	.5670	78.63	56.21
ACT mathematics	.0178	.2038	1.51	14.12
ACT composite	.1402	.3129	5.47	18.24
Exam score	.6794	.7342	68.63	73.21

The numerical coefficients in the normalized vector do not come directly from the discriminant equations. In the process of going from the discriminant equation to the normalized vector, the coefficients were adjusted, so that the sum of the squared coefficients was equal to 1.00. The values within the scaled vectors were compared with one another so as to determine which of the predictor variables were most effective as predictors within the context of the corresponding discriminant equation.

The values in the normalized vector could not be compared in this manner as they were highly analogous to the regression coefficients in a multiple regression prediction equation, whicle the values in the scaled vectors may be thought of a similar to beta weights.

Four discriminant equations were obtained for the groups; however, the first two were found to account for 88.5 per cent of the total discriminative power for the four variables used. Chi square tests (Bartlett, 1941) were computed for each of the two derived discriminate functions and each was found to be significant ($p < .001$). Table 3 shows the discriminating power of the functions as measured by the per cent of among-group variance explained.

Table 3

Discrimination Power of the Discriminant Functions

Discriminant Function	Per Cent of Total
I	71.7
II	16.8
III	9.4
IV	2.4

As the first two equations accounted for nearly 90 per cent of the total discriminative power, it was decided to use only those functions in the validation process. Figure 2 graphs the two functions based on the means found in Table 6. The centroids for each of the groups in Figure 2 were found by a simple two-step process. First, the means for a particular criterion group on each predictor were substituted for the appropriate *Zs* in the first discriminant equation. After multiplying the group means by the coefficients the result was labeledand plotted on the horizontal axis, in a similar manner, the same was done for the second discriminate equation and plotted on the vertical axis. This process was followed for each of the five groups.

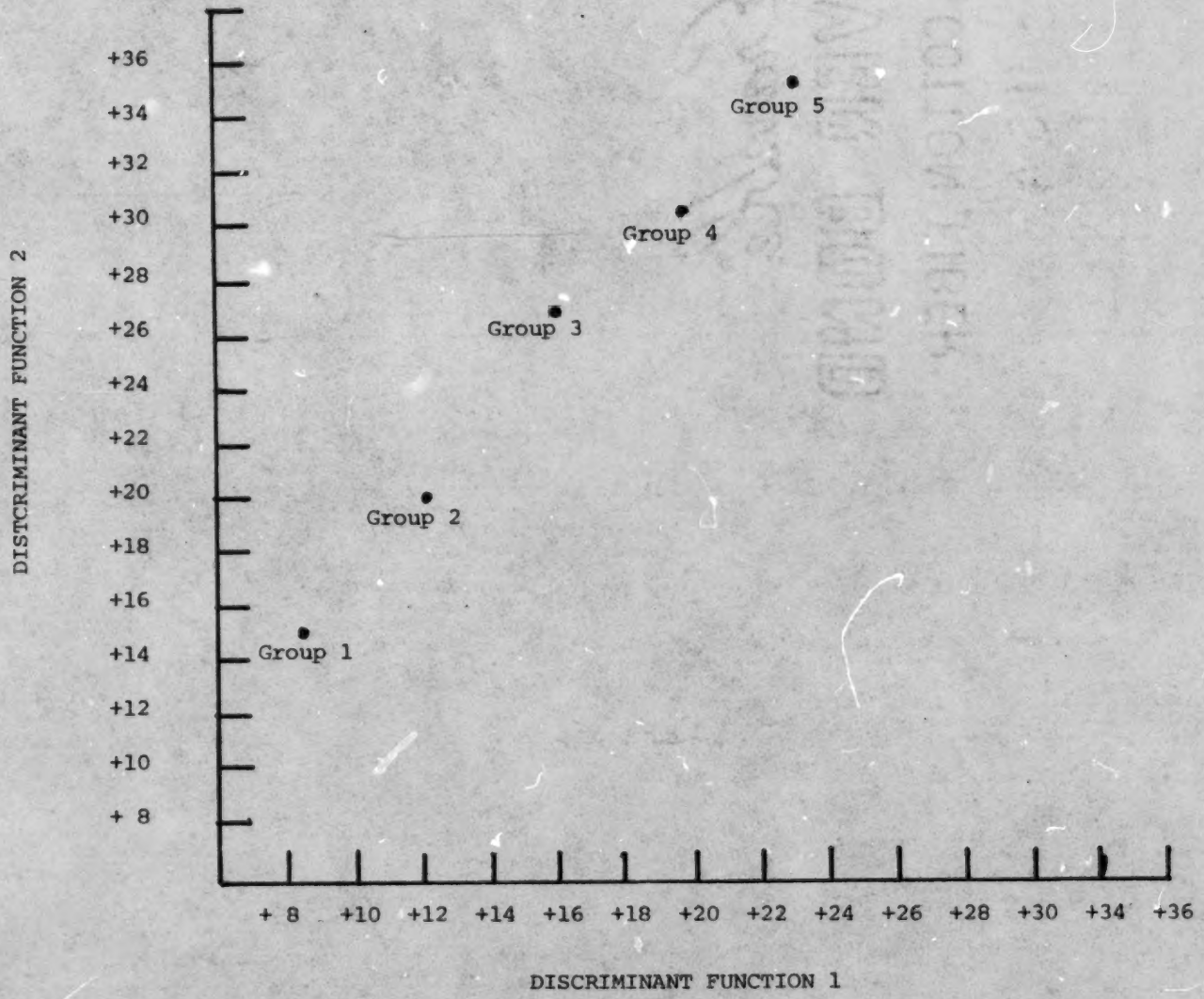
Table 6

Centroids of Groups in Discrminate Space

	FUNCTION	
GROUP	I	II
1	8.8	14.78
2	12.2	19.26
3	17.66	26.45
4	20.65	29.77
5	24.31	35.92

Figure 2

Graph of Group Centroids



To find out whether the prediction functions had a chance of being successful with a new group of individuals, cross-validation was performed between 1978 and 1979 freshmen. Using the 1978 freshmen, a separatediscriminate function equation was developed. This equation was used to predict a criterion score for people in the 1979 data set. The predicted criterion scores were then correlated with their actual criterion scores. The resulting correlation was .825 with a standard error of estimate of .239. An F-test of this multiple correlation revealed F = 4.22 which was significant at the .01 level. Because of the high correspondence between the two groups, it was concluded that cross-validation had been successful.

DISCUSSION

This study is still an on-going one and much further enhancements are anticipated. However, the prediction equations that were developed appear to discriminate among the groups effectively. The two predictor variables that discriminated the most effectively in the study were high school background and the grade on the departmental exam form.

In the very near future it is expected that not only will equations be developed to place students into courses; but, also, to indicate to the students what their expected chances are in terms success and failure in the course as measured on a four point scale.

Needless to say, the prediction equations are not stand alone devices; they must be given consideration only in light of the student's desires and the faculty member's recommendation.

REFERENCES

Tatsuoka, M. *Discriminant Analysis: The Study of Group Differences.* Champaign, Ill.: Institute for Personality and Ability Testing, 1970, pp.54-69.

_____. *Multivariate Analysis, Techniques for Educational and Psychological Research.* New York: John Wiley & Sons, 1971.

Wilks, S. "Certain generalizations in the analysis of variance." *Biometrika*, 1932, pp. 24-36.

Nie, *et al.* *Statistical Package for the Social Sciences, Second Edition.* New York: McGraw-Hill Book Company, 1975, pp. 434-467.